



## OPEN ACCESS

# Rating early child development outcome measurement tools for routine health programme use

Dorothy Boggs,<sup>1,2</sup> Kate M Milner,<sup>1,3</sup> Jaya Chandna,<sup>4</sup> Maureen Black,<sup>5,6</sup> Vanessa Cavallera,<sup>7</sup> Tarun Dua,<sup>7</sup> Guenther Fink,<sup>8</sup> Ashish KC,<sup>9</sup> Sally Grantham-McGregor,<sup>10</sup> Jena Hamadani,<sup>11</sup> Rob Hughes,<sup>12,13</sup> Karim Manji,<sup>14</sup> Dana Charles McCoy,<sup>15</sup> Cally Tann,<sup>1,16</sup> Joy E Lawn<sup>1</sup>

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/archdischild-2018-315431>).

For numbered affiliations see end of article.

## Correspondence to

Dorothy Boggs, London School of Hygiene & Tropical Medicine, London, UK; [dorothy.boggs@lshtm.ac.uk](mailto:dorothy.boggs@lshtm.ac.uk)

Received 1 November 2018

Revised 23 January 2019

Accepted 28 January 2019

## ABSTRACT

**Background** Identification of children at risk of developmental delay and/or impairment requires valid measurement of early child development (ECD). We systematically assess ECD measurement tools for accuracy and feasibility for use in routine services in low-income and middle-income countries (LMIC).

**Methods** Building on World Bank and peer-reviewed literature reviews, we identified available ECD measurement tools for children aged 0–3 years used in ≥1 LMIC and matrixed these according to *when* (child age) and *what* (ECD domains) they measure at population or individual level. Tools measuring <2 years and covering ≥3 developmental domains, including cognition, were rated for accuracy and feasibility criteria using a rating approach derived from Grading of Recommendations, Assessment, Development and Evaluations.

**Results** 61 tools were initially identified, 8% (n=5) population-level and 92% (n=56) individual-level screening or ability tests. Of these, 27 tools covering ≥3 domains beginning <2 years of age were selected for rating accuracy and feasibility. Recently developed population-level tools (n=2) rated highly overall, particularly in reliability, cultural adaptability, administration time and geographical uptake. Individual-level tool (n=25) ratings were variable, generally highest for reliability and lowest for accessibility, training, clinical relevance and geographical uptake.

**Conclusions and implications** Although multiple measurement tools exist, few are designed for multidomain ECD measurement in young children, especially in LMIC. No available tools rated strongly across *all* accuracy and feasibility criteria with accessibility, training requirements, clinical relevance and geographical uptake being poor for most tools. Further research is recommended to explore this gap in fit-for-purpose tools to monitor ECD in routine LMIC health services.

## BACKGROUND

The Sustainable Development Goals (SDGs) and *Global Strategy for Women's Children's and Adolescents' Health 2016–2030* envision a world where every child can survive and 'thrive', reaching their full developmental potential.<sup>1 2</sup> Global policy in early child development (ECD) is encapsulated within the WHO, UNICEF and World Bank Nurturing Care Framework (NCF)

## Key findings

1. **WHY? Multiple tools:** of the 100 tools that exist for early child development (ECD) outcome measurement, 27 met criteria for rating (measurement started <2 years and covered at least three developmental domains), however few are fit-for-purpose for use in routine health systems.
2. **WHAT IS NEW? Remit and range of tools:** of the tools identified, few adequately address multiple domains required for monitoring ECD, with the majority omitting vision and hearing. The two population-level tools rated highest in reliability, cultural adaptability, administration time and geographical uptake. The individual-level screening and ability tools rated highest for reliability and lowest for accessibility, training, clinical relevance and geographical uptake.
3. **WHAT TO DO? Accuracy and feasibility of tools:** few existing tools are both accurate (ie, valid, reliable) and feasible for training and routine use (eg, time, cost, accessibility) in LMIC settings.
4. **KEY GAPS?** The population-level tools (Caregiver Reported Early Development Instruments and Indicators of Infant and Young Child Development), along with the D-Score, are being harmonised into the WHO-led Global Scale for Early Development for population and programmatic level measurement. An optimal individual-level tool remains a gap. Additional research on tool assessment is needed to improve reporting, links to action and utility in planning and evaluating early intervention.

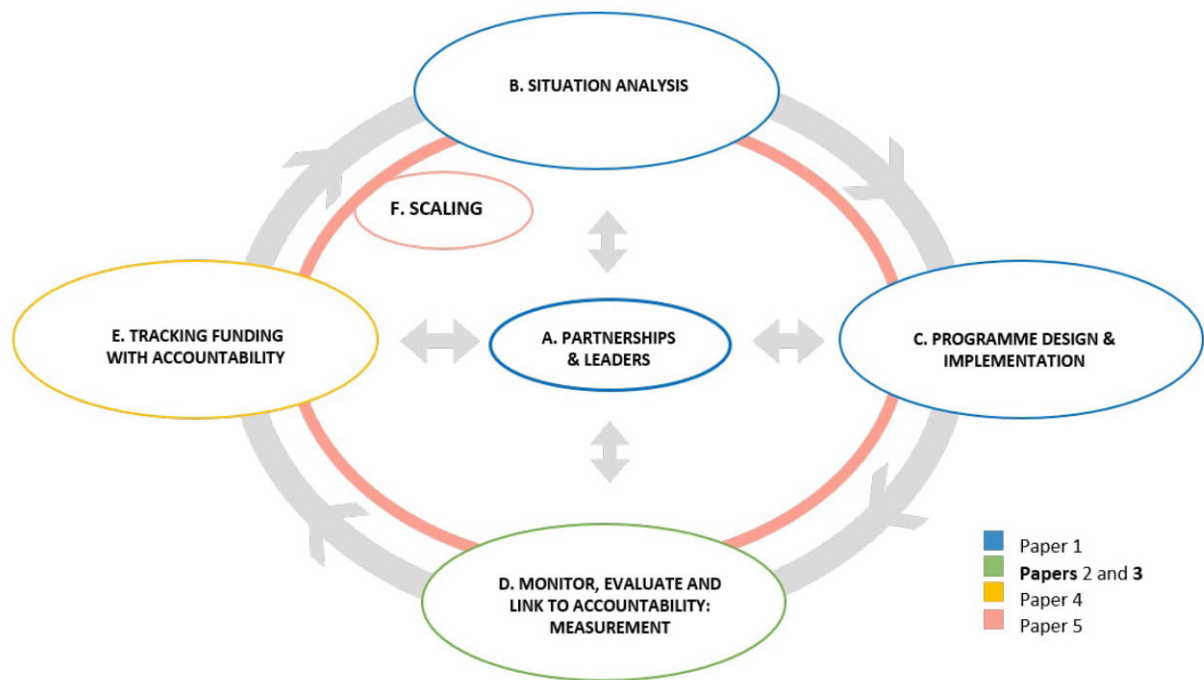
and low-income and middle-income countries (LMIC) have increasingly supported this 'beyond survival' agenda with 45% (68 countries) having national level ECD policies and programmes.<sup>3 4</sup>

Birth to 3 years is well-established as the critical period for ECD, when returns on investment are greatest.<sup>5–8</sup> Seizing this window requires early identification of children with developmental difficulties, particularly through existing large-scale maternal, newborn and child health (MNCH) programmes such as health surveillance immunisation and growth monitoring.<sup>3 9 10</sup>



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Boggs D, Milner KM, Chandna J, et al. *Arch Dis Child* 2019;**104**:S22–S33.



**Figure 1** Programme cycle for design, implementation and scaling of early child development programmes.

Developmental monitoring in high-income countries (HIC) has been shown to improve early identification and access to intervention for children at risk of developmental delay and/or impairment.<sup>11–14</sup>

As highlighted in this series, challenges exist in monitoring and evaluation of ECD programmes, and also for measurement of outcomes in routine systems, despite a plethora of tools.<sup>15</sup> Over the past several years, there have been several reviews for ECD measurement.<sup>16–20</sup> The most recent and comprehensive review, the World Bank's *Toolkit for Measuring Early Child Development in Low-income and Middle-income Countries*, provided an update to their previous toolkit and alongside published the *ECD Measurement Inventory* which summarised a total of 147 tools for children up to 8 years of age with reviews of peer-reviewed and grey literature up until 2017 (hereafter referred to as the World Bank's *Toolkit* and *Inventory*, respectively).<sup>16 19 21</sup>

In this paper, we systematically evaluate multidomain measurement of ECD in LMIC with a new and specific focus on those tools that measure a range of domains and could be applied for young children from 0 to 3 years of age through routine health services.

### Scope and structure of series

This paper is the third in a series examining evidence to inform design and implementation of ECD interventions at national and subnational level in LMIC. The series is structured around a programme cycle including key processes and decision points (figure 1). This paper focuses on potential ECD monitoring and evaluation tools for routine health services. Other papers have reviewed overall design decisions,<sup>9</sup> monitoring and evaluation,<sup>15</sup> financing<sup>22</sup> and overall process to scale-up.<sup>23</sup>

### AIM AND OBJECTIVES

We review ECD measurement tools for children 0–3 years of age and systematically assess appropriateness for use in routine health services in LMIC.

Our objectives are to:

1. Identify existing ECD measurement tools covering ages 0–3 years according to initial selection criteria (ie, including  $\geq 2$  domains, used in at least one LMIC).
2. Matrix these ECD measurement tools according to *when* (age) and *what* (domains) are included.
3. Rate accuracy and feasibility of selected tools that meet further eligibility criteria (commencing under 2 years of age and including  $\geq 3$  domains, one of which is cognition) according to a systematic rating approach for these tools characteristics.

### METHODS

#### Objective 1: identify existing ECD measurement tools covering ages 0–3 years

The World Bank's *Toolkit* and *Inventory* is the most comprehensive review of ECD measurement tools for use in LMIC.<sup>16 21</sup> The latest World Bank toolkit, published in December 2017, involved reviews of peer-reviewed literature regarding child development measurement tools in LMIC through keyword searches of PubMed, Google Scholar, PsycINFO and other databases, as well as grey literature including other collections.<sup>16–19 21 24</sup> We also reviewed recent reviews<sup>17 18 24</sup> and consulted experts including the coauthors on this paper to identify relevant tools not included in the *Inventory*. Tools were categorised according to purpose (population and individual levels) and type of measurement (ability, screening and both) as defined in web supplementary web appendix 1.

#### Objective 2: matrix ECD measurement tools according *when* (age) and *what* (domains) are included

A matrix was developed to cross tabulate *when* measurement is performed (child age bands) and *what* developmental domains

are measured. The age bands were based on the early years, considering likely opportunities for measurement within existing MNCH programmes in LMIC (eg, immunisation, growth monitoring). The domains were selected based on standard domains measured in global burden of disease assessments, which are also consistent in most clinical assessments.<sup>16 21 25</sup> We considered those domains used by the World Bank, such as motor, cognition, and others, notably vision and hearing.

From all the tools identified in objective 1, we mapped onto the matrix those tools which had been used in at least one LMIC as defined by the World Bank Country Income Groups and covered  $\geq 2$  developmental domains.<sup>26–29</sup> We used the cut-off  $\geq 2$  developmental domains as the standard clinical definition for global developmental delay or impairment.<sup>27–29</sup>

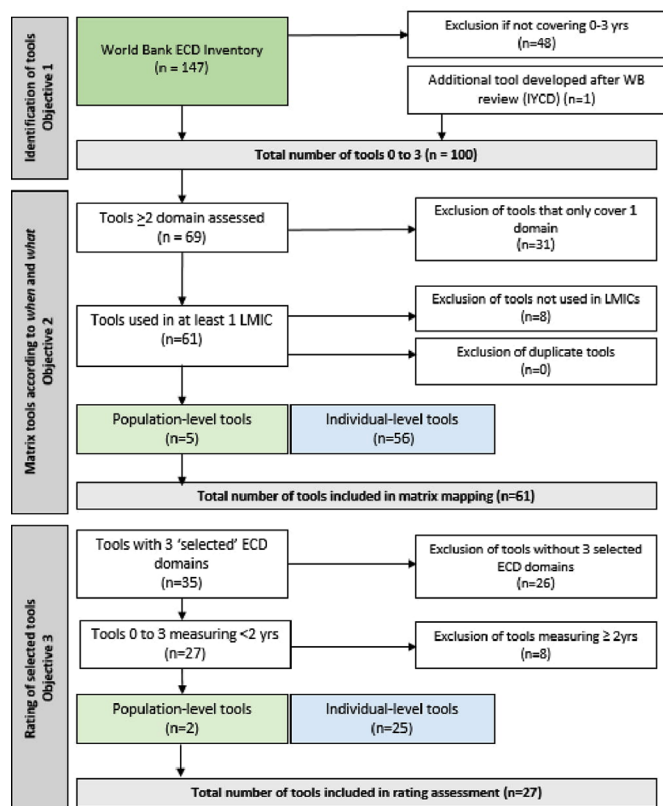
### Objective 3: rate accuracy and feasibility of selected tools

Tools measuring  $< 2$  years of age and including  $\geq 3$  domains, one of which is cognition, were selected for rating to ensure earlier multidomain measurement alongside health surveillance immunisation and growth monitoring.

Rating of tool characteristics focused on a minimum of seven distinct criteria, informed by the literature and agreed by the author group. Items focused on tool accuracy (ie, ‘Does the tool work?’) were informed by developmental measurement literature, including available existing literature focused on LMIC, and focused on validity, reliability and cultural adaptability.<sup>20 30 31</sup> Feasibility criteria (ie, ‘Can the tool be delivered?’), particularly informed by Fischer *et al*’s work, who assessed feasibility of ECD screening tools for use

**Table 1** Rating criteria for assessing early child development measurement tool accuracy and feasibility for use in routine programmes

Grading criteria	Definition	Rating	Meaning
A. Does the tool work? Psychometric properties and cultural adaptability of tool			
1. Validity	The degree to which a measure accurately assesses behaviours or abilities that reflect the underlying concept being tested. (16)	3	Validity ideally against educational outcomes up to age 5 with a standardised test, eg, Wechsler, equal to or above widely accepted threshold (eg, >0.7), statistically significant.
		2	Validity somewhat below widely accepted threshold (eg, 0.5–0.7) against another performance-based tool, eg, Bayley III.
		1	Some description/mention of validity but methods unclear or poor quality, below accepted threshold (eg, <0.5).
		0	Inadequate result of validity, no statistical significance.
2. Reliability	How consistently a measure produces similar results for a child or group of children with repeated measurements over a short period of time. (16)	3	Equal to or above widely accepted threshold (eg, >0.7) for measure tested at tool level, rigorous methods of testing, statistically significant ideally with kappa. (supplementary web appendix 1).
		2	Somewhat below widely accepted threshold (eg, 0.5–0.7), rigorous methods of testing but in one continent only.
		1	Some description/mention but methods unclear or poor quality or below accepted threshold (eg, <0.5).
		0	Inadequate discussion of reliability, no statistical.
3. Cultural adaptability	Modification of items, materials and procedures to fit the local context, such as translating items and changing words or pictures to reflect cultural differences. (16)	3	Easy modification of items, materials and procedures.
		2	Minimum to moderate modification of items, materials and procedures.
		1	Moderate to complex modification of items, materials and procedures.
		0	Highly difficult modification of items, materials and procedures.
B. Can the tool be delivered? Practicality of administration			
1. Accessibility	Access to tool, including digital availability and costs to purchase and use the tool with equipment as required. <i>Note: cost is allocated per child for 100 tests.</i> <i>Note: digital defined here as open access tool available online and app available.</i> <i>Note: cost does not include training costs, some tools may be freely available but require payment for a trainer to train the project team.</i>	3	Tool, administration, scoring and interpretation, adaptation and training resources all available open access online with no intellectual property restrictions; no cost for tool, no additional equipment; app available.
		2	Tool, administration, scoring and interpretation, adaptation and training resources all available open access online with no intellectual property restrictions, minimal cost to tool and/or equipment (≤US\$10 per child), no app available.
		1	Tool, administration, scoring and interpretation, adaptation and training resources all available online, but some intellectual property or other restrictions (eg, requirement for direct involvement tool authors/owners in research), moderate cost to tool and/or equipment (range >US\$10 to ≤US\$20 per child), no app available.
		0	Not readily available online with intellectual property restrictions, high cost tool and equipment (range >US\$20 per child), no app available.
2. Training	Refers to duration of training, skill level of trainer and trainee and certification requirement. <i>Note: duration of training does not include general field work.</i>	3	Brief (≤1 hour), minimal (ie, non-specialist worker can train non-specialist worker), no certification requirement.
		2	Moderate (>1 hour to ≤1 day), moderate (ie, non-specialist trainer) but requires more standardisation and training or direct assessments of children’s abilities that require moderate training and practice, no certification requirement.
		1	Long (≤2 days), moderate (ie, non-specialist trainer) but requires more standardisation and training or direct assessments of children’s abilities that require moderate training and practice, may include certification requirement.
		0	Long (≥3 days), specialist trainer and/or trainee, certification required.
3. Administration time	Estimated time taken to administer the tool in completion, including scoring time. <i>Note: when range is given an estimated median time for administration will be used.</i>	3	≤15 min, easy scoring.
		2	>15 to ≤30 min, minimum to moderate scoring.
		1	>30 to ≤60 min, moderate to complex scoring.
		0	>60 min.
4. Geographical uptake	Geographical use of the tool.	3	Used in at least three continents.
		2	Used in two continents only.
		1	Used in one continent only.
		0	Used in one country only.
C. Individual-level tool only			
1. Clinical relevance and utility	Usability of tool for frontline worker for interpretation and response.	3	Easy interpretation, clear threshold for action and structure for counselling response and contextually appropriate referral.
		2	Minimum to moderate interpretation, thresholds for action but unstructured response guidance and/or suggested response unlikely to be feasible in context.
		1	Moderate to complex interpretation, no structured thresholds for action and/or suggested response unfeasible in context.
		0	Highly technical interpretation (eg, with separate manual), no clear threshold for action, specialist referral response



**Figure 2** Early child development (ECD) tools flow chart for multidomain matrix mapping and grading. IYCD, Infant and Young Child Development; LMIC, low- and middle-income countries.

by community health workers in LMIC, focused on tool accessibility, training, administration time and geographical uptake<sup>18 20 29</sup>. An eighth criteria for clinical relevance and utility was included for individual-level tools only since population-level tools are not intended to measure individual-level assessment. Rating criteria for assessing early child development measurement tool accuracy and feasibility for use in routine programmes is presented in [table 1](#) (online supplementary appendix 1).

Rating of tools for each of these characteristics was informed by the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) system as a guiding framework.<sup>32</sup> The GRADE system is widely used including by WHO and scores four levels of evidence quality (high, moderate, low and very low) and classifies recommendations as strong or weak.<sup>33 34</sup> Hence for each of the characteristics in our rating tool, we consistently applied a descending scale of '3' to '0' (ie, four levels), according to strength and/or utility of each tool characteristic based on available evidence.

To apply the rating, two authors independently rated the tools (agreement 92%) and consensus was reached with KMM by reviewing evidence for ratings which were not in agreement. To identify evidence for rating each tool, one of the authors searched on PubMed, Google Scholar databases and on Google for easily available peer-reviewed and grey literature, including individual tool manuals and test websites (supplementary web appendix 2).

In view of the potential bias introduced by excluding tools for not measuring the 'cognitive' domain (as per the World Bank *Inventory*'s definition), we also rated four excluded tools, analysing at least one tool from each of the three groups (population, or individual screening or ability).

## RESULTS

### Objective 1: identify existing ECD measurement tools covering ages 0–3 years

Out of the 147 tools included in World Bank's *Inventory*, 99 tools covered the ages 0–3 years ([figure 2](#) and supplementary web appendix 3).<sup>21</sup> The WHO Indicators of Infant and Young Child Development (IYCD), released after the World Bank's review, was identified separately by experts.<sup>35–37</sup>

### Objective 2: matrix ECD measurement tools according to when (age) and what (domains) are included

#### Matrix development

Our matrix included the following age intervals: 3 months to first year, then 6 months until 3 years of age ([figure 3](#)). Two additional age groups of children aged >3 years were also included for ongoing developmental monitoring. Nine developmental domains, as defined in [table 2](#) according to the World Bank's *Inventory*, were included, including two learning domains, which are typically tested from 2.5 years onwards and noting that some domains were more constructs.<sup>16 21</sup> Hearing and vision, particularly critical given their importance in broader development and lack of universal screening for sensory impairments in LMIC, were also included for a total of 11 domains.<sup>16–18 24 38 39</sup>

#### Tool mapping

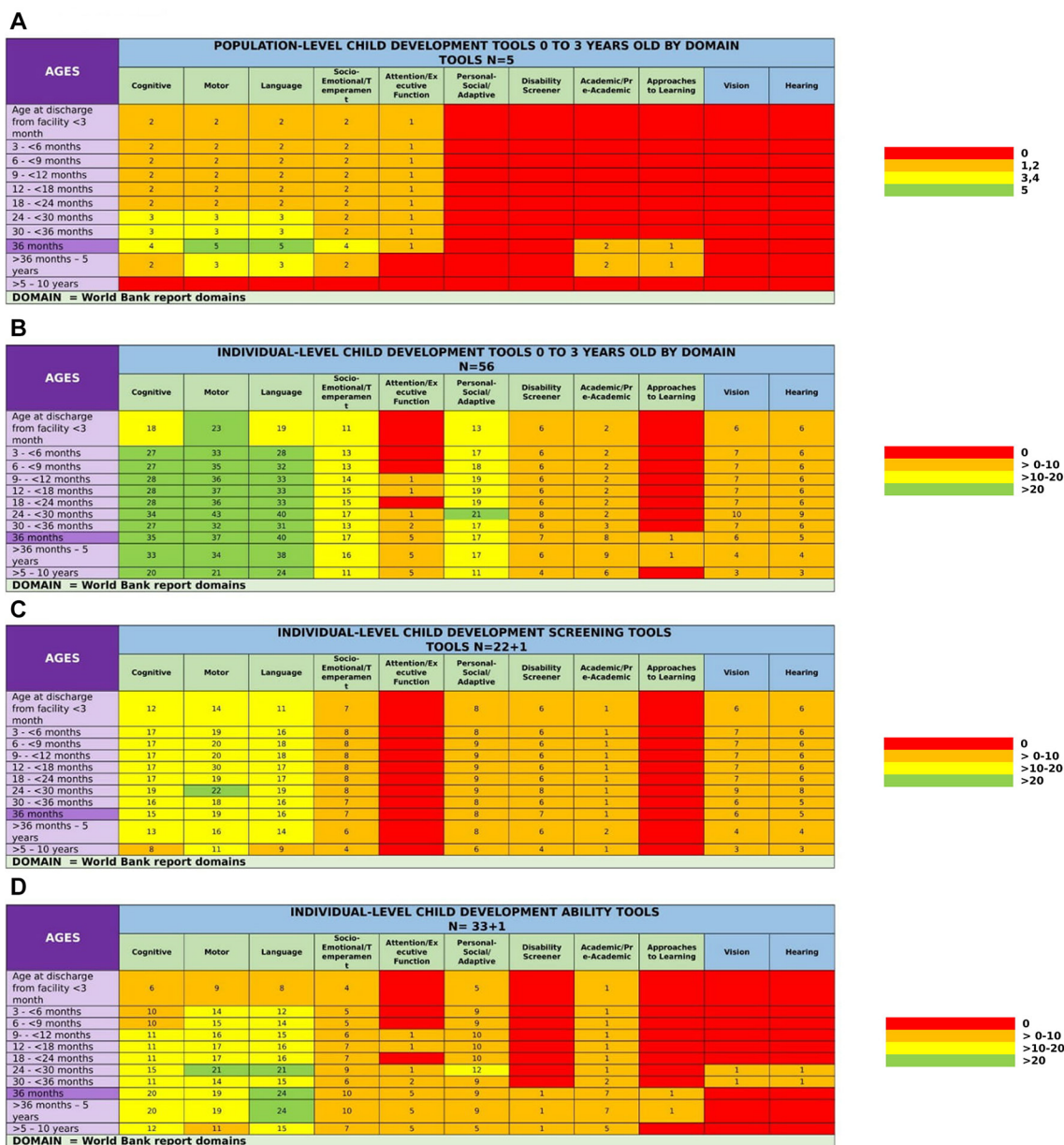
Sixty-one tools met criteria for inclusion onto the matrix. The majority of tools (92%, n=56) were individual-level screening (n=22) or ability tests (n=33), while the remaining 8% (n=5) were population-level tools ([figure 2](#)). One tool, the Movement Assessment Battery for Children, was identified as an individual-level screening and ability test so is counted in both categories ([figure 2](#) and supplementary web appendix 3).<sup>40</sup>

Cognitive, motor and language domains were most commonly included across all tool groups. At the population-level, 60% (n=3) tools included these three domains from 24 to 36 months, and all five tools (100%) measured motor and language at 36 months. At the individual-level (n=56), 55% (n=31) of tools measured all three domains. Specifically, in screening tools (n=23), the motor domain was measured from 24 to <30 months of age in 96% (n=22) tools. In ability tools (n=34), both motor and language was measured from 24 to <30 months in 62% (n=21) tools.

There were noticeable measurement gaps in most other domains for all tool types, especially in ages 0–3 years. No population-level tool covered personal-social adaptive, disability screener, vision and hearing domains ([figure 3A](#)). In individual-level tools, there were noticeable gaps in measurement of attention/executive function, disability, academic preacademic, approaches to learning, vision and hearing domains, with fewer than 25% (n=12) measuring each domain from 0 to 3 years ([figure 3B](#)). No screening tool measured attention/executive function or approaches to learning, and fewer than 40% (n=9) measured each socioemotional/temperament, personal-social/adaptive, disability screener, academic preacademic, vision and hearing domain across 0–3 years ([figure 3C](#)). Less than 36% (n=12) of ability tools measured each of the remaining eight domains from 0 to 3 years, with disability, approaches to learning, vision or hearing only measured by one ability tool each across a very limited age range ([figure 3D](#)).

Individual tool mappings can be found in online web appendix 4.





**Figure 3** Heat map matrix of early child development measurement tools 0–3 inclusion of identified times, ages and domains. (A) Population-level. (B) Individual-level (screening, ability and both screening and ability tools). (C) Screening tools. (D) Ability tools.

### Objective 3: rate accuracy and feasibility of selected tools

Forty-eight per cent (n=27) of tools met criteria for inclusion for rating of accuracy and feasibility (figure 2 and supplementary web appendix 3). Total ratings (figure 4) were analysed for the 27 tools for each characteristic and tool with recommendations classified as strong to weak (figures 5 and 6).<sup>32–34</sup>

#### Population-level tools (n=2)

Two population-level tools rated strongly for both accuracy and feasibility criteria, with high ratings in cultural adaptability as well as in accessibility, administration time and geographical uptake.<sup>35 37 41 42</sup> Caregiver-Reported Early Child Development Instruments (CREDI) rated strongest within the population-level tools, rating strongly in validity and reliability in well documented multicountry studies and moderately for training.<sup>41</sup> The IYCD rated very low in training, low in validity and moderate in

reliability, as the tool's complete psychometric results are forthcoming.<sup>35 37 42</sup>

#### Individual-level screening tools (n=14)

These demonstrated great variability, rating between 0 and 20. The Guide for Monitoring Child Development rated strongest within the individual-level tools, followed by Parents' Evaluation of Developmental Status (PEDS) and then Ages and Stages Questionnaire (ASQ).<sup>43–45</sup> The Developmental Screening Questionnaire rated lowest, with all characteristics rating either very low (n=2) or not known (n=6).<sup>46</sup> Overall, this tool group had the strongest ratings for administration time and strongest ratings for reliability with 50% (n=7) rating strongly (ie, 3) for this characteristic. Accessibility was 'not known' or 'very low' for 71% (n=10) and geographical uptake was also 'very low' for 29% (n=4). 50% (n=7) of tools in this group rated

**Table 2** Nine developmental domains up to 3 years of age from the World Bank's *Inventory*<sup>16</sup>

Domains	Definition
1 Cognitive	The test assesses cognitive development, including general intellectual ability, problem-solving, conceptual development, reasoning, visual-spatial ability, memory, learning, etc.
2 Language	The test assesses language development/ability, including receptive and/or expressive language.
3 Motor	The test assesses motor development/ability including fine and/or gross motor.
4 Socioemotional/temperament	The test assesses socioemotional development or temperament, which are overlapping constructs, especially in the early years. Socioemotional development includes behaviour problems, social competency, emotional competency and self-regulation. Temperament includes extraversion/surgency (positive affect, activity level, impulsivity, risk-taking), negative affectivity (fear, anger, sadness, discomfort) and effortful control (attention shifting and focusing, perceptual sensitivity, inhibitory and activation control).
5 Attention/executive function	The test assesses executive function, including attention, working memory, inhibitory control, cognitive flexibility, planning, etc.
6 Personal-social/adaptive	The test assesses personal-social or adaptive skills or self-help skills, such as feeding, dressing, toilet training, recognising and interacting with others.
7 Academic/preacademic	The test assesses academic or preacademic skills, such as literacy and math/numeracy.
8 Approaches to learning	The test assesses approaches to learning.
9 Disability screener	The test was designed to screen children for disability or severe developmental delay.

'not known' for cultural adaptability and clinical relevance and utility.

#### Individual-level ability tools (n=11)

Ratings for ability tests also varied widely (ie, ratings 3–16) with Intergrowth 21st Neurodevelopment Assessment (INTER-NDA) rating highest and The Oxford Neurodevelopment Assessment (OX-NDA) rating lowest.<sup>47 48</sup> Overall, this tool group rated highest on psychometrics including reliability and then validity, although fewer than 20% (n=2 and n=1, respectively) were rated as 'strong' in each characteristic. 55% (n=6) rated 'not known' for cultural adaptability and 'very low' in accessibility, training and geographical uptake. 73% (n=8) rated either 'very low' or 'not known' for clinical relevance and utility.

#### Rating of tools excluded in World Bank document

To address potential exclusion bias, four tools that had been excluded since they did not formally measure 'cognition' as per the World Bank's *Inventory* were rated (box 1). The Malawi Developmental Assessment Tool (MDAT) rated strongly tied with INTER-NDA for highest rating of 16 for the individual-level ability tools.

## DISCUSSION

This paper systematically rates the accuracy and feasibility of multidomain ECD 0–3 measurement tools with an explicit focus on routine use within the health sector in LMIC. Despite a plethora of ECD tools, our results indicate that none cover all domains and are accurate and feasible.<sup>24</sup> Among the 27 ECD tools that were rated, no tool adequately covered the majority

of the domains or rated strongly for *all* accuracy and feasibility grading characteristics. However, at least one tool rated highly enough in each group: CREDI for population-level tools, GMCD for individual-level screening tools followed closely by PEDS and ASQ, and INTER-NDA for individual-level ability tools. These results have important implications for ECD measurement within health programmes by identifying existing tools that can be used and are reliable yet more feasible, for example, requiring shorter administration time or less complex training.

Cognitive, language and motor domains were most frequently measured, with gaps across other domains. Vision, hearing and disability screener were missing in all in population-level tools, along with the personal-social/adaptive domain, and <20% (n=9/56) individual-level tools measured these domains. Vision, hearing and disability screening are critical at population-level and individual-level for early identification of developmental delay and/or impairment and to ensure referrals and/or follow-up for children identified. The academic/preacademic and approaches to learning domains, typically measured from age 2.5 years onwards, were perhaps understandably not frequently assessed given our aged-restricted inclusion criteria as well as higher level attention and executive functions.

Overall, accuracy characteristics were most difficult to obtain information on rating, with validity evidence rarely detailed. Generally, all three tool groups rated more strongly in reliability than validity, with 10 tools rating a '3' for reliability and only three tools rating a '3' for validity. More research is required to better test and document psychometric properties in LMIC, in order to meet more rigorous validity criteria, such as 'strong' which is to be predictive validity in different contexts.<sup>28</sup> Since using HIC norms is not optimal, tools need to have a local comparison group of reference or control children for standardisation.<sup>28 49</sup> Furthermore, a noticeable documentation gap in accuracy characteristics was cultural adaptability, with half of all tools rated as 'not known'. Often studies cited which items were modified during translation/back translation but did not discuss the process and/or the complexity of implementing this process.<sup>44 46 48 50–62</sup> An example of good documentation is mentioned in the study by Gladstone *et al*, which detailed the adaptation process of creating a culturally relevant developmental assessment tool in rural Africa.<sup>63</sup> In future, accuracy should be reported in a more standardised way and the adaptation process better documented.<sup>30 64 65</sup>

Feasibility information was easier to locate compared with other criteria, although ratings were typically lower. Administration time and geographical uptake characteristics rated highest across all tools and both were well documented in the World Bank *Inventory*, although those authors acknowledged that the country list was not exhaustive.<sup>21</sup> No tool rated strongly on training criteria indicating a need to aim for shorter tool trainings by non-specialist trainers.

Only the two population-level tools CREDI and IYCD rated strongly for accessibility, indicating the majority of ECD tools are not readily and freely accessible online with app availability for use, highlighting another area for future improvement.

Almost half of all individual-level tools rated clinical relevance as 'not known' and only a quarter of individual-level tools rated this highly; considering clinical relevance is usually the basis for referral and follow-up, this highlights a critical gap for front-line workers. It is important this criteria is easy to interpret with clear thresholds for action and structured counselling responses, and especially essential that accessible service infrastructure for assessment of children who screen positively needs to be in place.

ECD TEST	RATING CRITERIA								TOTAL RATING
	Does the tool work? Psychometric properties and cultural adaptability of tool			Can the tool be delivered? Practicality of administration				Individual-level tool only	
	Validity	Reliability	Cultural adaptability	Accessibility	Training	Administration time	Geographical uptake	Clinical relevance and utility	
	Degree to which a measure accurately assesses behaviours or abilities that reflect the underlying concept being tested	How consistently a measure produces similar results for a child or group of children with repeated measurements over a short period of time	Modification of items, materials, and procedures to fit local context, such as translating items and changing words or pictures to reflect cultural differences	Access to tool, including digital availability and costs to purchase and use the tool with equipment as required	Refers to duration of training, skill level of trainer and trainee, and certification requirement	Estimated time taken to administer the tool in completion, including scoring time	Geographical use of the tool	Usability of tool for front line worker for interpretation and response	
POPULATION-LEVEL									MAX 21
Caregiver-Reported Early Child Development Instruments (CREDI)	3	3	3	3	2	3	3		20
Indicators of Infant and Young Child Development (IYCD)	1	2	3	3	0	3	3		15
<b>TOTAL RATING</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>6</b>	<b>2</b>	<b>6</b>	<b>6</b>		
INDIVIDUAL-LEVEL									MAX 24
Screening tests n=14									
Abbreviated Developmental Scale I (EAD-1)	1	2	NK	NK	2	2	0	NK	7
Access Portfolio	1	NK	NK	NK	2	2	2	3	10
Ages & Stages Questionnaire (ASQ)	2	3	2	0	2	3	3	3	18
Ages & Stages Questionnaire: Inventory	NK	NK	NK	NK	NK	NK	2	NK	2
Ankara Developmental Screening Inventory	2	3	NK	1	2	1	2	NK	11
Child Development Evaluation (CDE)	2	1	0	2	2	3	0	NK	10
Denver Developmental Screening Test	1	3	2	0	0	3	3	2	14
Denver Prescreening Developmental Questionnaire (Denver II PDQ)	1	3	1	0	2	2	3	2	14
Developmental Screening Questionnaire (DSQ)	NK	NK	NK	0	NK	NK	0	NK	0
Guide for Monitoring Child Development (GMCD)	3	3	3	0	2	3	3	3	20
ICMR Psychosocial Development Screening Test	NK	3	NK	2	2	3	1	NK	11
Parents' Evaluation of Developmental Status (PEDS)	2	3	2	2	2	3	2	3	19
Rapid Neurodevelopmental Assessment Tool: 0-2 y	2	2	2	0	0	2	2	2	12
Shoklo Neurological and Developmental Tests	2	0	NK	NK	1	2	0	NK	5
<b>TOTAL RATING</b>	<b>19</b>	<b>26</b>	<b>12</b>	<b>7</b>	<b>17</b>	<b>29</b>	<b>23</b>	<b>18</b>	
ABILITY TESTS N=11									
Battelle's Development Inventory	2	2	NK	1	0	0	0	3	8
Bayley Scales of Infant Development (BSID-III)	3	2	1	0	0	1	3	1	11
Comprehensive Developmental Inventory for Infants and Toddlers (CDIIT)	2	2	NK	0	0	0	0	NK	4
Developmental Assessment Scale for Indian Infants (DASII)	2	2	NK	0	0	1	0	2	7
Early Childhood Care and Development (ECCD) Checklist	NK	NK	NK	2	1	0	0	3	6
Extended Ages & Stages Questionnaire (EASQ)	NK	1	1	0	2	NK	3	NK	7
Griffiths Mental Development Scales	2	3	2	0	0	1	3	0	11
Intergrowth 21st Neurodevelopment Assessment (INTER-NDA)	2	3	2	2	1	3	3	NK	16
Mullen Scales of Early Learning (MSEL)	2	2	2	0	0	1	2	0	9
Test de Aprendizaje y Desarrollo Infantil (TADI)	2	2	NK	1	1	2	0	NK	8
The Oxford Neurodevelopment Assessment (OX-NDA)	NK	NK	NK	NK	1	2	0	NK	3
<b>TOTAL RATING</b>	<b>17</b>	<b>19</b>	<b>8</b>	<b>6</b>	<b>6</b>	<b>11</b>	<b>14</b>	<b>9</b>	

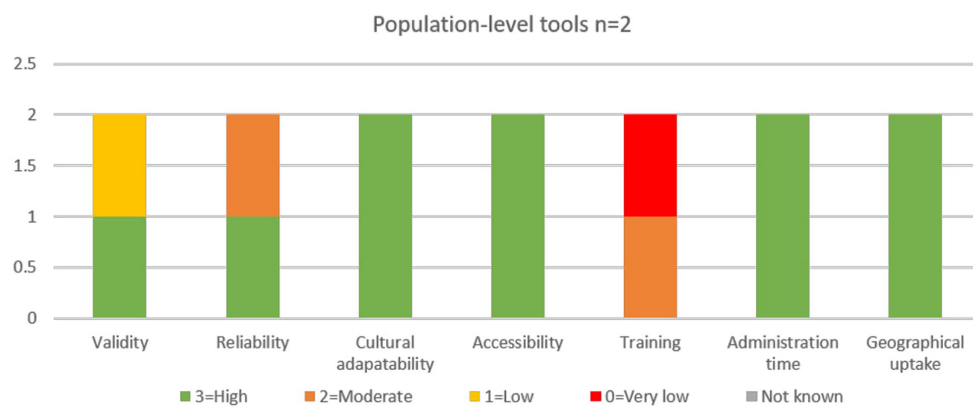
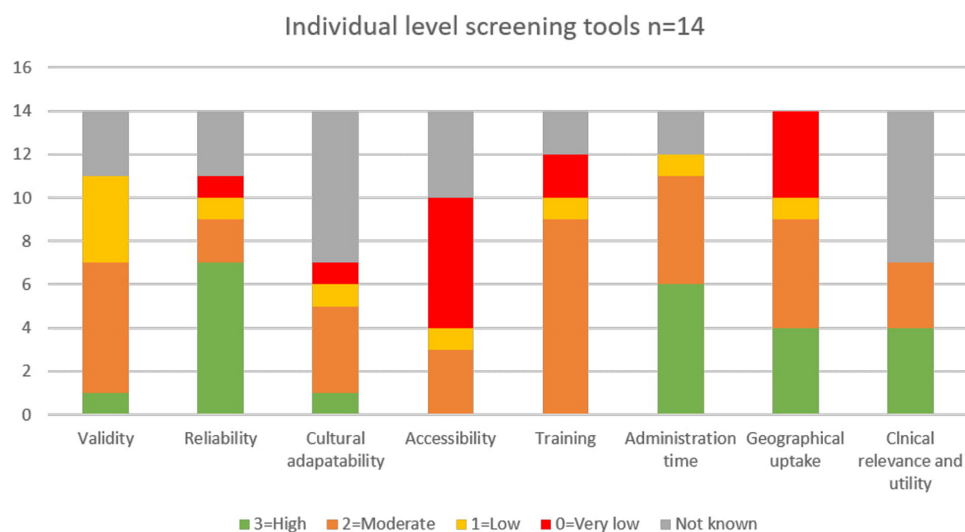
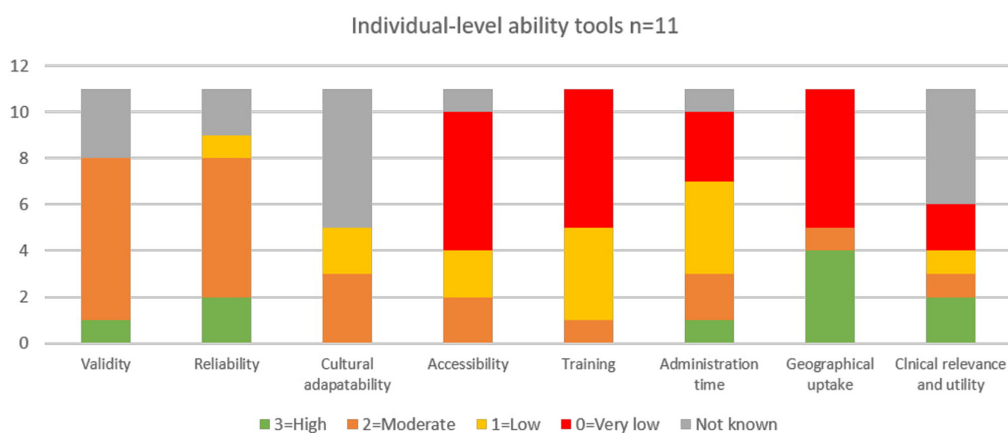


Figure 4 Heat map of accuracy and feasibility ratings for selected early child development (ECD) measurement tools.

For comparison, the review by Fischer *et al* recommended Ten Questions Questionnaire (TQQ), GMCD and MDAT for feasibility of use in LMIC health settings. Although our analysis rated GMCD highly, TQQ and MDAT were excluded for grading as they respectively do not measure development for children <2 years or formally document measuring cognition as per World Bank's definition. Box 1 shows results of rating four tools which were excluded, where MDAT rated 16, tied with INTER-NDA for highest individual-level ability tool rate, indicating a limitation of the filter definitions.

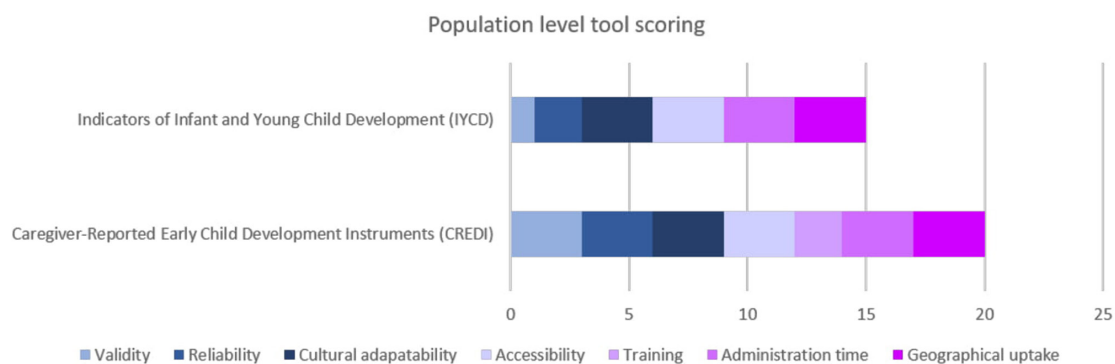
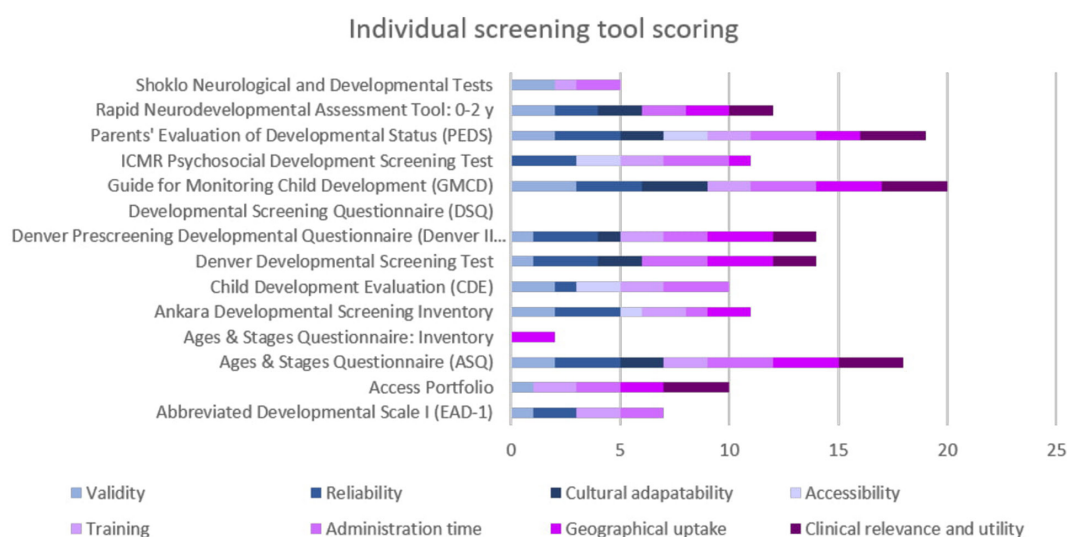
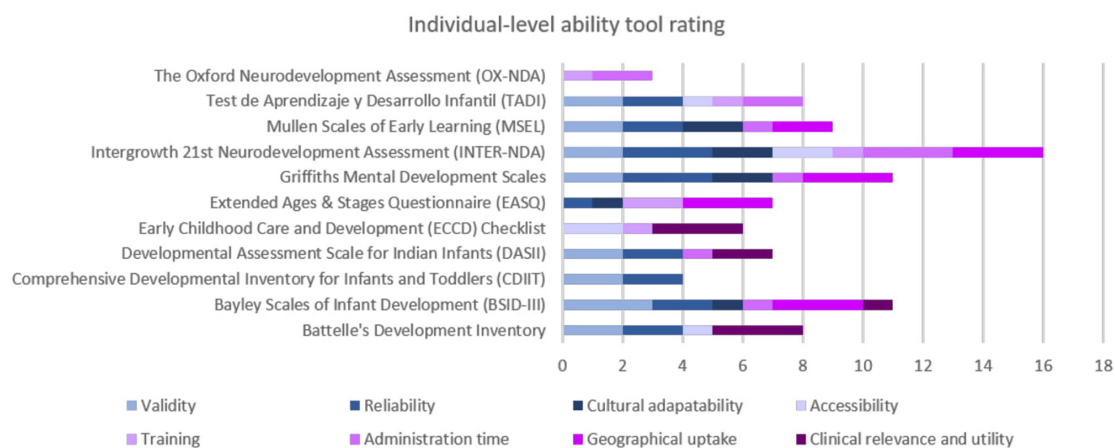
### Strengths and limitations

The World Bank's *Toolkit* and *Inventory* were recently published and provided crucial input for our work, and identified 106 new tools for a total of 147 ECD tools 0–8 years.<sup>16 21</sup> However, this framing may also have had limitations. For example, filters based on the *Inventory* provided a useful way to categorise tool content; however, this categorisation also limited analysis of domains, such as personal-social/adaptive which could be measured through other domains, and of other tools, such as those that did not adhere to the specific 'cognitive' domain

**a) Population-level tools****b) Individual-level screening tools****c) Individual-level ability tools**

**Figure 5** Rating criteria characteristic heat map for early child development tools 0–3 years. (A) Population-level tools. (B) Individual-level screening tools. (C) Individual-level ability tools.



**a) Population-level tools****b) Individual-level screening tools****c) Individual-level ability tools**

**Figure 6** Early child development 0–3 tool overall rating mapped by each accuracy and feasibility criteria. (A) Population-level tools. (B) Individual-level screening tools. (C) Individual-level ability tools.

## Box 1 Consideration of tools excluding the World Bank's 'cognitive' domain

In the World Bank's *inventory*, the cognitive domain was defined as 'the test assesses cognitive development, including general intellectual ability, problem-solving, conceptual development, reasoning, visual-spatial ability, memory, learning, etc'.<sup>16 21</sup> Although tools were usefully categorised as 'yes' if they explicitly measured this domain, other tools were categorised as 'no' despite measuring cognition implicitly alongside other child development domains. This was often due to the child development tool measuring aspects of cognition, but not listing it formally as one of the formal domains measured.

On review of the tools that were excluded when the three selected cognitive, language and motor domain filter was applied in objective 3 (online supplementary web appendix 3), it was noted that many of these tools do in fact measure cognition. Given this finding, four tools were selected across the three tool categories (population-level, individual-level screening and individual-level ability) for rating to address this possible exclusion bias to compare these rates with the 27 tools that were initially rated. This methodology followed a similar method as outlined in the main paper, except KMM was the second reviewer for MDAT.

The ratings are shown below:

	Validity	Reliability	Cultural adaptability	Accessibility	Training	Administration time	Geographical uptake	Clinical relevance and utility	
<b>ADDITIONAL TOOLS</b>									<b>MAX 21</b>
Multiple Indicator Cluster Surveys (MICS) Early Child Development Index (ECDI)	NK	NK	0	2	2	3	3		10
Screening Test Battery for Assessment of Psychosocial Development	NK	3	NK	0	NK	2	0	NK	5
Malawi Developmental Assessment Tool (MDAT)	2	3	2	2	0	2	2	3	16
Kilifi Developmental Checklist (KDC)	2	3	NK	0	1	3	0	NK	9

All four of these tools demonstrated good rating potential with evidence available.

MICS ECDI, the population-level tool, rated a '10' which is lower than the CREDI and IYCD tool rates of 20 and 15, respectively. MICS ECDI rated strongly on administration and geographical uptake, however has noticeable psychometric gap with validity and reliability unknown. The Screening Test Battery for Assessment of Psychosocial Development, the individual-level screening tool, rated a low rate of 5. This tool rated strongly in reliability, which was consistent with this tool group, however either rated 'very low' or 'not known' in six of the eight tool characteristics.

However, it is the individual-level ability tool category which is most notably striking. The MDAT rated highest in this supplemental analysis with a rate of 16, which is tied for the highest overall rate in this tool category with INTER-NDA and the KDC rated 9, which is more similar to other tools in this category (figure 4). MDAT rated moderately or strongly in seven of the tool characteristics, and evidence was available for all criteria. It is also noted that MDAT covers a much broader age range 0–8 years, compared with INTER-NDA 22–26 months.

Although the recent World Bank's *Toolkit* and *Inventory* have advanced the ECD field, this finding indicates that caution might need to be applied when applying the filters with their respective definitions for further analysis.

MICS, Multiple Indicator Cluster Surveys ECDI, Early Child Development Index; INTER-NDA, Intergrowth 21st Neurodevelopment Assessment; KDC, Kilifi Developmental Checklist; MDAT, Malawi Developmental Assessment Tool; NK, not known.

definition (box 1). When imposing filters the tools' 'country used' information is not exhaustive, and the vision and hearing domains may not be comprehensive. Newer, lesser known tools and those not available in English, or used in one country may also be under-represented, as well as specific tools measuring multidomain disabilities or impairments in young children. This was due to the World Bank's primary focus on ECD 0–8 years and less for early identification of children with multidomain disabilities or impairments; however, it is important to note that most LMIC cannot afford separate screening systems.

Tools with the highest rates were generally more widely used with more documentation in the public domain; hence these higher rates might reflect increased use as much as, or more than, accuracy and feasibility. Also, although some tools rated low on certain criteria, it is acknowledged that they may be suitable to purposes beyond health.

Finally, this review prioritised looking at ECD multidomain tools in young children; however, it is acknowledged that home context is extremely important alongside this measurement. As highlighted in the paper by Milner *et al*, contextual tools that measure both

maternal/caregiver mental health as well as caregiver capabilities, caregiver-child interactions and/or the home environment and long-term educational outcomes need to be considered.<sup>9 66–68</sup>

### Further research

This exercise highlights that ECD tool characteristics are inconsistently reported in literature and overall rated weakly on accuracy and feasibility characteristics. Following the development of the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement in response to inadequate reporting of observational studies, several extensions to STROBE have been created to provide more nuanced field-specific guidance for authors, such as for newborn infection.<sup>69 70</sup> Development of a STROBE extension checklist could establish standards for reporting on ECD tools and core data.<sup>69</sup> A systematic way to document these characteristics would reduce such inconsistencies. One step could be to expand on the *Quick checklist for appraising a CDAT* by Sabanathan *et al*, which provided five key questions for consideration of an assessment tool.<sup>24</sup>

In addition, the tools could be further examined according to administration of test (ie, caregiver report vs direct child observation), and this ECD tool mapping and rating exercises demonstrated both the strengths and limitations of employing a 'domain lens' approach to ECD tools. Although this provided a helpful classification for measurement and analysis for the purposes of this paper, it also highlights the need for widely agreed and established definitions for each domain, and the limits of imposing a siloed perspective and approach across tools (box 1), especially when domain specificity is not strong in infants and young children. Therefore, it is recommended that the ECD sector look more holistically at the child's functioning and environment when assessing and measuring children's abilities through health and other intersectoral areas such as education. Examples of moving away from a siloed domain perspective are exhibited in UNICEF and Washington Group on Disability Statistics' *Child functioning module* from 2 to 17 years of age, which assesses functional difficulties for censuses and national surveys,<sup>71–75</sup> and in recent review by Oberklaid *et al*, which highlights a move away from universal developmental surveillance using structured tools towards broader conversations and support with families in HIC.<sup>76</sup>

Finally, given the large number of tools available, there is a need for fit-for-purpose population-level and individual-level screening and ability tools that could better meet accuracy and feasibility criteria to monitor ECD in routine LMIC health services. Joint work at the population-level is currently in process. The CREDI and IYCD teams have come together with the Global Child Development Group (developers of the Developmental 'D-Score' Growth Chart) to form the Global Scales for Early Development (GSED).<sup>36</sup> The GSED will include a single set of open-access metrics for capturing population-level ECD for children under 3 years, as well as a programme evaluation measure. As part of this process, this group is considering many of the issues outlined above, including the reliability, validity, cross-cultural applicability and feasibility, greatly enhancing ECD measurement and monitoring at population-level. Following on from this work, an individual-level fit-for-purpose tool is equally needed for both global screening and ability testing purposes. It is recommended that these approaches are aligned and adhere to a similar process, especially giving key consideration to the accuracy and feasibility criteria.

## CONCLUSION

Improved measurement of ECD in routine maternal, newborn and child health services is urgently needed to ensure that programme implementation and monitoring are aligned with *The Global Strategy* and the SDG targets, especially in terms of reaching the most vulnerable young children at highest risk of developmental delays and/or impairment. Despite multiple tools exist for measuring ECD outcomes in children aged 0–3 years, few adequately meet accuracy and feasibility criteria for use at either population or individual levels. Recently developed population-level child development measurement tools are promising, but further research is required to develop accurate and feasible individual-level tools for use in routine health programmes at scale in LMIC. In addition, more consistent reporting of studies of the development and use of ECD tools is necessary to allow comparisons and more rapid learning.

## Author affiliations

<sup>1</sup>Maternal, Adolescent, Reproductive and Child Health Centre, London School of Hygiene and Tropical Medicine, London, UK

<sup>2</sup>International Centre for Evidence in Disability, London School of Hygiene and Tropical Medicine, London, UK

<sup>3</sup>Murdoch Children's Research Institute, Melbourne, Victoria, Australia

<sup>4</sup>Institute of Translational Medicine, University of Liverpool, Liverpool, UK

<sup>5</sup>University of Maryland School of Medicine, Baltimore, Maryland, USA

<sup>6</sup>Research Triangle Park, RIT International, Durham, USA

<sup>7</sup>Department of Mental Health and Substance Abuse, World Health Organisation, Geneva, Switzerland

<sup>8</sup>Swiss Tropical and Public Health Institute and University of Basel, Basel, Switzerland

<sup>9</sup>International Maternal and Child Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

<sup>10</sup>Institute of Child Health, Faculty of Population Health Sciences, University College London, London, UK

<sup>11</sup>Maternal and Child Health Division, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh

<sup>12</sup>Children's Investment Fund Foundation, London, UK

<sup>13</sup>Maternal & Child Health Intervention Research Group, Department of Population Health, London School of Hygiene and Tropical Medicine, London, UK

<sup>14</sup>Department of Paediatrics and Child Health, Muhimbili University of Allied Health Sciences, Dar es Salaam, Tanzania

<sup>15</sup>Harvard Graduate School of Education, Harvard University, Massachusetts, USA

<sup>16</sup>Neonatal Medicine, University College Hospitals NHS Trust, London, UK

**Acknowledgements** The authors would like to thank all the World Bank team for the extensive and helpful review. The authors would like to thank Dr Melissa Gladstone for her contributions and review of this paper. The authors would also like to thank Victoria Ponce Hardy for compiling and formatting figures and references, and to Claudia da Silva and Fion Hay for administrative assistance.

**Contributors** Technical oversight of the series was led by JEL and KMM. The first draft of the paper and analysis was undertaken by DB, with input from KMM and JEL. JC was the second scorer. The Early Child Development Expert Advisory Group (Pia Britto, TD, Esther Goh, SG-McG, MG, JH, RH, KMM, Jamie Radner, Muneera Rasheed, Karlee Silver, Arjun Upadhyay) contributed to the conceptual process throughout. All authors gave input to scoring criteria and reviewed the manuscript.

**Funding** This supplement has been made possible by funding support from the Bernard van Leer Foundation. Saving Brains® impact and process evaluation funded by Grand Challenges Canada®.

**Disclaimer** The authors alone are responsible for the views expressed in this article and they do not necessarily represent the views, decisions or policies of the institution with which they are affiliated.

**Competing interests** The following authors on this paper have intellectual inputs and leadership roles for some of the tools reviewed: MDAT (JC), IYCD (VC, TD) and CREDI (DCM and GF). None of these authors rated any of these tools.

**Patient consent for publication** Not required.

**Provenance and peer review** Commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- United Nations. Sustainable Development Goals, 2015.
- Every Woman Every Child. *The Global Strategy for Women's, Children's and Adolescents' Health (2016–2030)*, 2015.
- Richter LM, Daelmans B, Lombardi J, *et al*. Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *Lancet* 2017;389:103–18.
- World Health Organisation. *Nurturing Care Framework*. Geneva: World Health Organisation, 2018. [https://www.who.int/maternal\\_child\\_adolescent/documents/nurturing-care-early-childhood-development/en/](https://www.who.int/maternal_child_adolescent/documents/nurturing-care-early-childhood-development/en/)
- Walker SP, Wachs TD, Grantham-McGregor S, *et al*. Inequality in early childhood: risk and protective factors for early child development. *Lancet* 2011;378:1325–38.
- Institute of Medicine. In: Shonkoff J, Phillips A, eds. *From neurons to neighbourhoods. The science of early childhood development*. Washington, DC: The National Academies Press, 2000.
- World Health Organisation. *A Critical Link. Interventions for physical growth and psychological development Department of Child and Adolescent Health and Development*, World Health Organisation, 1999.
- Heckman J. The case for investing in disadvantaged young children. *CESifo DICE Report* 2008;6:3–8.
- Milner KM, Bernal R, Brentani A, *et al*. Contextual design choices and partnerships for scaling early child development programmes. *Arch Dis Child* 2019;104(Suppl 1):S3–S12.
- Dua T, Tomlinson M, Tablante E, *et al*. Global research priorities to accelerate early child development in the sustainable development era. *Lancet Glob Health* 2016;4:e887–e889.



- 11 Guevara JP, Gerdes M, Localio R, *et al.* Effectiveness of developmental screening in an urban setting. *Pediatrics* 2013;131:30–7.
- 12 van Agt HM, van der Stege HA, de Ridder-Sluijer H, *et al.* A cluster-randomized trial of screening for language delay in toddlers: effects on school performance and language development at age 8. *Pediatrics* 2007;120:1317–25.
- 13 Oberklaid F, Efron D. Developmental delay-identification and management. *Aust Fam Physician* 2005;34:739–42.
- 14 Moore TG, McDonald M, Carlon L, *et al.* *Early childhood development and the social determinants of health inequities: A review of the evidence.* Victoria, Australia: Victorian Health Promotion Foundation, 2015;30:ii102–ii115.
- 15 Milner KM, Bhopal S, Dua T, *et al.* Counting outcomes, coverage and quality for early child development programmes. *Arch Dis Child* 2019;104(Suppl 1):S13–S21.
- 16 Fernald L, Prado E, Kariger P, *et al.* *A toolkit for measuring early childhood development in low- and middle-income countries.* Washington DC: International Bank for Reconstruction and Development/The World Bank, 2017.
- 17 Semrud-Clikeman M, Romero RAA, Prado EL, *et al.* [Formula: see text]Selecting measures for the neurodevelopmental assessment of children in low- and middle-income countries. *Child Neuropsychol* 2017;23:1–42.
- 18 Fischer VJ, Morris J, Martinez J, *et al.* Developmental screening tools: feasibility of use at primary healthcare level in low- and middle-income settings. *J Health Popul Nutr* 2014;32:314–26.
- 19 Fernald L, Kariger P, Engle P, *et al.* *Examining Early Child Development in Low-Income Countries: A Toolkit for the Assessment of Children in the First Five Years of Life.* Washington DC: World Bank, 2009.
- 20 Rubio-Codina M, Araujo MC, Attanasio O, *et al.* Concurrent Validity and Feasibility of Short Tests Currently Used to Measure Early Childhood Development in Large Scale Studies. *PLoS One* 2016;11:e0160962.
- 21 *Early Child Development Measurement Inventory [Internet]:* World Bank, 2017. (cited 2018).
- 22 Arregoces L, Hughes R, Tann C, *et al.* Accountability for funds for Nurturing Care: what can we measure? *Arch Dis Child* 2019;104(Suppl 1):S34–S42.
- 23 Cavallera V, Tomlinson M, Radner J, *et al.* Scaling early child development: what are the barriers and enablers? *Arch Dis Child* 2019;104(Suppl 1):S43–S50.
- 24 Sabanathan S, Wills B, Gladstone M. Child development assessment tools in low-income and middle-income countries: how can we use them more appropriately? *Arch Dis Child* 2015;100:482–8.
- 25 Institute for Health Metrics and Evaluation. IHME: Measuring what matters, 2018. Available: <http://www.healthdata.org/gbd/>
- 26 The World Bank Group. Country Income Groups (World Bank Classification), World Bank Country and Lending Groups. 2018 <http://data.worldbank.org/about/country-classifications/country-and-lending-groups> (30 Aug 2018).
- 27 Shevell M, Ashwal S, Donley D, *et al.* Practice parameter: Evaluation of the child with global developmental delay: Report of the quality standards subcommittee of the American academy of neurology and the practice committee of the child neurology Society. *Neurology* 2003;60:367–80.
- 28 Majnemer A, Shevell MI. Diagnostic yield of the neurologic assessment of the developmentally delayed child. *J Pediatr* 1995;127:193–9.
- 29 Mithyantha R, Kneen R, McCann E, *et al.* Current evidence-based recommendations on investigating children with global developmental delay. *Arch Dis Child* 2017;102:1071–6.
- 30 Glascoe F, Cairney J. Best practices in test construction for developmental-behavioral measures: quality standards for reviewers and researchers. In: N H, BJ J, eds. *In follow-up for ncu graduates - promoting positive developmental and behavioral outcomes for at-risk infants.* New York, New York: Springer Publishing, 2018.
- 31 American Education Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for education and psychological testing:* American Psychological Association, 2014.
- 32 Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- 33 Guyatt GH, Oxman AD, Kunz R, *et al.* Going from evidence to recommendations. *BMJ* 2008;336:1049–51.
- 34 Guyatt GH, Oxman AD, Kunz R, *et al.* What is “quality of evidence” and why is it important to clinicians? *BMJ* 2008;336:995–8.
- 35 World Health Organisation. *Infant and young child development report.*
- 36 McCoy D, Black M, Daelmans B, *et al.* Measuring child development in children from birth to age 3 at population level. *Early Childhood Matters* 2016.
- 37 Lancaster GA, McCray G, Kariger P, *et al.* Creation of the WHO Indicators of Infant and Young Child Development (IYCD): metadata synthesis across 10 countries. *BMJ Glob Health* 2018;3:e000747.
- 38 Vohr B. Long-term outcomes of moderately preterm, late preterm, and early term infants. *Clin Perinatol* 2013;40:739–51.
- 39 Gogate P, Gilbert C, Zin A. Severe visual impairment and blindness in infants: causes and opportunities for control. *Middle East Afr J Ophthalmol* 2011;18:109–14.
- 40 Henderson S, Sudgen DB A. Movement assessment battery for children: Pearson. 2007. <https://www.pearsonclinical.com/therapy/products/100000433/movement-assessment-battery-for-children-second-edition-movement-abc-2.html>
- 41 Harvard TH. Chan School of Public Health. Caregiver-Reported Early Development Instruments (CREDI). 2018. <https://sites.sph.harvard.edu/credi/>
- 42 Infant and Young Child Development IYCD Package. <https://ezcollab.who.int/iycd>
- 43 PEDS Test. Available: <http://www.pedstest.com/default.aspx>
- 44 Ages & Stages Questionnaires. Translations of ASQ. <https://agesandstages.com/languages/>
- 45 Ertem I. The international guide for monitoring child development: enabling individualised interventions. *Early Childhood Matters* 2017.
- 46 Kwan C, Nam S. Utilizing parental observations and computer technology in developing a child-screening instrument in Singapore. *Int J Early Years Educ* 2004;12:117–29.
- 47 The Inter-NDA Consortium. Assessing early child development. <https://www.inter-nda.com/about.html>
- 48 INTERGROWTH-21st. *The INTERGROWTH-21st Neurodevelopmental Assessment (INTER-NDA) Manual:* University of Oxford, 2013.
- 49 Bodeau-Livinec F, Davidson LL, Zoumenou R, *et al.* Neurocognitive testing in West African children 3–6 years of age: Challenges and implications for data analyses. *Brain Res Bull* 2019;145:129–35.
- 50 Muñoz-Caicedo A, Zapata-Osso H, Perez-Tenorio L. Validacion de criterio de la Escala Abreviada del Desarrollo (EAD-1) en el dominio audicion-lenguaje. *Revista de Salud Publica* 2013;15.
- 51 Wirz S, Edwards K, Flower J, *et al.* Field testing of the ACCESS materials: a portfolio of materials to assist health workers to identify children with disabilities and offer simple advice to mothers. *Int J Rehabil Res* 2005;28:293–302.
- 52 Sezgin N. Two Different Validity Study of Ankara Developmental Screening Inventory (ADSI): Criterion-Related Validity and Concurrent Discrimination Validity. *Turk J Child Adolesc Ment Health* 2011;18:185–96.
- 53 Tsai YP, Tung LC, Lee YC, *et al.* Selecting score types for longitudinal evaluations: the responsiveness of the Comprehensive Developmental Inventory for Infants and Toddlers in children with developmental disabilities. *Neuropsychiatr Dis Treat* 2016;12:1103–9.
- 54 Centre MR. Developmental Assessment Scales for Indian Infants (DASII). <https://www.manashakti.org/tests/developmental-assessment-scales-indian-infants>
- 55 *How to use the early childhood care and development checklist.* In: Council E, ed.
- 56 Early Childhood Care and Development (ECCD) Checklist, Child's Record 2.
- 57 Glascoe FP, Byrne KE. The usefulness of the Battelle Developmental Inventory Screening Test. *Clin Pediatr* 1993;32:273–80.
- 58 CEDEP. Test de aprendizaje y desarrollo infantil. 2013. <http://www.cedep.info/tadi.html>
- 59 Vazir S, Landsdown R, Naidu A, *et al.* A comparison of Indian and American Scales of Child Development. *Journal of the Indian Academy of Applied Psychology* 1994;20:175–81.
- 60 Haataja L, McGready R, Arunjerda R, *et al.* A new approach for neurological evaluation of infants in resource-poor settings. *Ann Trop Paediatr* 2002;22:355–68.
- 61 Mukherjee SB, Aneja S, Krishnamurthy V, *et al.* Incorporating developmental screening and surveillance of young children in office practice. *Indian Pediatr* 2014;51:627–35.
- 62 Indian council of medical research. <https://icmr.nic.in/>
- 63 Gladstone MJ, Lancaster GA, Jones AP, *et al.* Can Western developmental screening tools be modified for use in a rural Malawian setting? *Arch Dis Child* 2008;93:23–9.
- 64 Borman J, Sevcik RA, Romski M, *et al.* Successfully translating language and culture when adapting assessment measures. *J Policy Pract Intellect Disabil* 2010;7:111–8.
- 65 Peña ED. Lost in translation: methodological considerations in cross-cultural research. *Child Dev* 2007;78:1255–64.
- 66 Caldwell B, Bradley RH. *Home Observation for Measurement of the Environment: Administration Manual.* Tempe, AZ: Family & Human Dynamics Research Institute: Arizona State University, 2003.
- 67 Kariger P, Frongillo EA, Engle P, *et al.* Indicators of family care for development for use in multicountry surveys. *J Health Popul Nutr* 2012;30:472–86.
- 68 Patient Health Questionnaire (PHQ-9 & PHQ-2): American Psychological Assessment. <http://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/patient-health.aspx>
- 69 Sharp MK, Utrobičić A, Gómez G, *et al.* The STROBE extensions: protocol for a qualitative assessment of content and a survey of endorsement. *BMJ Open* 2017;7:e019043.
- 70 Fitchett EJA, Seale AC, Vergnano S, *et al.* Strengthening the reporting of observational studies in epidemiology for newborn infection (STROBE-NI): an extension of the STROBE statement for neonatal infection research. *Lancet Infect Dis* 2016;16:e202–e213.
- 71 Loeb M, Cappa C, Criallesi R, *et al.* Measuring child functioning: the Unicef/ Washington Group Module. *Salud Publica Mex* 2017;59:485–7.
- 72 UNICEF. Module on Child Functioning concept note. 2017. <https://data.unicef.org/resources/module-child-functioning-concept-note/>
- 73 Washington Group on Disability Statistics. Child Functioning. <http://www.washingtongroup-disability.com/washington-group-question-sets/child-disability/>
- 74 World Health Organisation. *International Classification of Functioning, Disability and Health. ICF.* Geneva, Switzerland: World Health Organisation, 2001.
- 75 World Health Organisation. *The International Classification of Functioning, Disability and Health for Children and Youth. ICF-CY.* Geneva, Switzerland: World Health Organisation, 2007.
- 76 Oberklaid F, Baird G, Blair M, *et al.* Children's health and development: approaches to early identification and intervention. *Arch Dis Child* 2013;98:1008–11.